



King's Research Portal

DOI:

[10.18632/oncotarget.14140](https://doi.org/10.18632/oncotarget.14140)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Puig-Butille, J. A., Gimenez-Xavier, P., Visconti, A., Nsengimana, J., Garcia-García, F., Tell-Marti, G., Escamez, M. J., Newton-Bishop, J., Bataille, V., Río, M. D., Dopazo, J., Falchi, M., & Puig, S. (2017). Genomic expression differences between cutaneous cells from red hair color individuals and black hair color individuals based on bioinformatic analysis. *Oncotarget*, 8(7), 11589-11599. <https://doi.org/10.18632/oncotarget.14140>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Genomic expression differences between cutaneous cells from red hair color individuals and black hair color individuals based on bioinformatic analysis

Puig-Butille Joan Anton¹, Gimenez-Xavier Pol², Visconti Alessia³, Nsengimana Jérémie⁴, Garcia-García Francisco⁵, Tell-Marti Gemma², Maria José Escamez⁶, Newton-Bishop Julia⁴, Bataille Veronique³, del Río Marcela⁶, Dopazo Joaquín^{5,7,8}, Falchi Mario³, Puig Susana²

¹Biochemistry and Molecular Genetics Department, Melanoma Unit, Hospital Clinic & IDIBAPS, CIBER de Enfermedades Raras (CIBERER), Barcelona, Spain

²Dermatology Department, Melanoma Unit, Hospital Clinic & IDIBAPS, CIBER de Enfermedades Raras (CIBERER), Barcelona, Spain

³Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

⁴Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK

⁵Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

⁶Departamento de Bioingeniería, Universidad Carlos III de Madrid, CIEMAT, IIS-Fundación Jiménez Díaz, CIBER de Enfermedades Raras (CIBERER), Madrid, Spain

⁷Functional Genomics Node, (INB) at CIPF, Valencia, Spain

⁸CIBER de Enfermedades Raras (CIBERER), Valencia, Spain

Correspondence to: Joan Anton Puig-Butille, **email:** jantonpuig@gmail.com; japuig@clinic.ub.es

Keywords: MC1R, red hair phenotype, melanocyte, melanoma, autophagy

Received: July 28, 2016

Accepted: November 21, 2016

Published: December 24, 2016

ABSTRACT

The *MC1R* gene plays a crucial role in pigmentation synthesis. Loss-of-function *MC1R* variants, which impair protein function, are associated with red hair color (RHC) phenotype and increased skin cancer risk. Cultured cutaneous cells bearing loss-of-function *MC1R* variants show a distinct gene expression profile compared to wild-type *MC1R* cultured cutaneous cells. We analysed the gene signature associated with RHC co-cultured melanocytes and keratinocytes by Protein-Protein interaction (PPI) network analysis to identify genes related with non-functional *MC1R* variants. From two detected networks, we selected 23 nodes as hub genes based on topological parameters. Differential expression of hub genes was then evaluated in healthy skin biopsies from RHC and black hair color (BHC) individuals. We also compared gene expression in melanoma tumors from individuals with RHC versus BHC. Gene expression in normal skin from RHC cutaneous cells showed dysregulation in 8 out of 23 hub genes (*CLN3*, *ATG10*, *WIPI2*, *SNX2*, *GABARAPL2*, *YWHA*, *PCNA* and *GBAS*). Hub genes did not differ between melanoma tumors in RHC versus BHC individuals. The study suggests that healthy skin cells from RHC individuals present a constitutive genomic deregulation associated with the red hair phenotype and identify novel genes involved in melanocyte biology.

INTRODUCTION

Human cutaneous pigmentation is dependent on melanin pigment production (eumelanin and pheomelanin) by epidermal and follicular melanocytes.

Melanin synthesis is controlled by the melanocortin receptor type 1 (*MC1R*) which encodes a 7-pass transmembrane G-protein-coupled receptor. In wild-type *MC1R* melanocytes, activation of the receptor by the α -melanocyte stimulating hormone (α -MSH) promotes

the synthesis of eumelanin pigment (dark pigment), reduces UV-induced oxidative stress and enhances DNA repair through base-excision repair and NER mechanisms which repairs UV-photoproducts [1]. *MC1R* is a highly polymorphic gene and loss-of-function variants (p.R151C, p.R142H, p.R161W, p.D294H p.D84E and p.I155T) result in a minimal receptor activity and mainly produces pheomelanin (red/yellow) pigment [2].

Epidemiological studies have found that loss-of-function variants in *MC1R* can, in part, predict the red-hair color (RHC) phenotype (red hair, fair skin, low tan capacity and high UV sensitivity) as well as melanoma [3] and non-melanoma skin cancer risk [4]. The decreased eumelanin production compared to pheomelanin in RHC individuals increases skin cancer risk due to the weak UV shielding capacity of pheomelanin, increase in UVA-induced reactive oxygen species [5, 6] and altered NER pathway. However, recent data indicate that RHC *MC1R* variants also contribute to carcinogenesis by UV-independent mechanisms. Mitra D et al. observed that the absence of pheomelanin is protective against melanoma development in mice models [7]. They detected high levels of oxidative DNA and lipid damage in RHC mice in a UV-independent model which still leads to oxidative damage. In addition, we have previously reported that co-cultured melanocytes and keratinocytes harbouring loss-of-function *MC1R* variants show a constitutive overexpression of genes involved in oxidative phosphorylation pathway and DNA repair mechanisms [8].

The aim of this study was to further identify genes related to the RHC phenotype by analysing the previously reported expression signature pattern associated with RHC *MC1R* variants [8] using protein-protein interaction (PPI) network analyses. Gene co-expression networks have the potential to highlight specific molecular mechanisms and genes related to a specific trait or disease [9–11]. Expression of candidate genes from the PPI network analyses were further analysed in independent gene expression data generated from normal skin biopsy from RHC individuals and Black hair color (BHC) using the UK MuTHER dataset and in a subset of melanomas from RHC and BHC patients.

RESULTS

We identified 3,570 differentially expressed genes (DEGs) in co-cultured melanocytes and keratinocytes from a pair of RHC siblings carrying non-functional *MC1R* variants versus a pair of siblings with wild-type *MC1R* alleles as previously described [8]. The expression signature pattern associated with RHC *MC1R* variants was evaluated by PPI network analysis. Networks from up-regulated and down-regulated DEGs were constructed separately as, from a systems biology

perspective, functionally-related genes are frequently co-expressed across a set of samples [12] and up-regulated and down-regulated transcripts tend to sub-connect [13, 14]. A statistically significant network composed of 557 nodes was detected among the set of 1954 up-regulated genes ($p < 0.001$; Figure 1). Also, we found a statistically significant network composed of 450 nodes among the set of 1616 down-regulated genes ($p < 0.001$, Figure 2). Topological information of networks and name of nodes (genes) are indicated in Supplementary Data Table S1 and Table S2.

We evaluated degree and betweenness centrality in both networks to investigate relationship between nodes and select the hub genes in each network. Twenty-three nodes were selected as hub genes (Table 1). In both networks, highly connected nodes tend to also show high betweenness centrality values. However, two nodes with a low degree centrality value (*BUB1B* and *SNX2* gene) were selected based on their high Betweenness centrality value.

In the network of up-regulated genes, 13 nodes were classified as hub genes (Figure 3). These genes play a role in DNA repair and cell-cycle homeostasis (*PRKAA1*, *CDK1*, *BUB1B*, *PCNA*, *RPA1* and *BRCA1*), oxidative phosphorylation (*GBAS*, *ICT1*) and autophagy (*PIK3C3*, *ATG4C*, *ATG10* and *SNX2*). The neighbouring genes of each hub gene are listed in Supplementary Tables S3-S14.

In the network of down-regulated genes, 11 nodes were classified as hub genes (Figure 4). These genes are involved in apoptosis (*SMAD3*, *YWHAG*), mRNA metabolism (*PABPC1*), and autophagy (*TRAF2*, *SQSTM1*, *CLN3*, *WIPI2*, *GABARAPL1*, *GABARAPL2*, *MAP1LC3B*, *MAP1LC3A*). The neighbouring genes of each hub gene are listed in Supplementary Tables S15-S25.

Next, we assessed the differential expression of those 23 genes in an additional whole genome expression dataset of healthy skin biopsies from the TwinsUK MuTHER dataset in 14 RHC individuals and 7 BHC individuals. At 10% FDR, genome wide differential expression analyses identified 1,952 DEGs between RHC and BHC individuals, consisting of 1,074 up-regulated and 878 down-regulated genes in RHC (Supplementary Table S26). Overall, 378 DEGs were common between the healthy skin and the co-cultured melanocyte-keratinocyte datasets. No statistically significant PPI networks were detected among DEGs sets (data not shown). Expression data from 18 out of 23 hub genes identified in the melanocyte-keratinocyte network analysis was available in the MuTHER dataset (*MAP1LC3B* was not included in the array and *PABPC1*, *ATG4C*, *CDK1* and *PIK3C3* genes failed the quality control). Five out of 18 genes classified as hub genes in the melanocyte-keratinocyte network analysis were differentially expressed in the MuTHER dataset with RHC skin showing down-regulation of *CLN3* and *ATG10* genes and up-regulation of *SNX2*, *PCNA* and *GBAS* genes. When the differential expression analysis was restricted to these 18 hub genes, up-regulation of

GABARAPL2 and *YWHA* genes and down-regulation of *WPI2* gene also reached statistically significant values (Table 2).

Finally, expression of hub genes was evaluated in a whole genome expression dataset of 26 melanomas from 6 BHC and 20 RHC patients. None of these genes were found to be differentially expressed between tumors from RHC patients compared with BHC patients. Interestingly, the top two upregulated genes in melanomas from RHC were *PRKAA1* (fold change=1.37, $p=0.08$) and *PIK3C3* (fold change=1.23, $p=0.09$) while the top two downregulated were *YWHA* (fold change=0.62, $p=0.29$) and *CLN3* (fold change=0.77, $p=0.3$). These p -values were unadjusted for multiple testing.

DISCUSSION

In this study, we aimed to uncover genes associated with the RHC phenotype by analysing a previously reported gene expression pattern [8]. By PPI network analysis, we identified two gene co-expression networks that reached statistically significant values in loss-of-function *MC1R* variants cutaneous cells. Since highly connected nodes are central to the network's architecture [15], the study was focused on those nodes with higher degree and betweenness centrality values. Overall, 23 nodes were selected as hub genes and their expression was analysed in an additional dataset from non-lesional skin tissue from RHC individuals or BHC individuals.

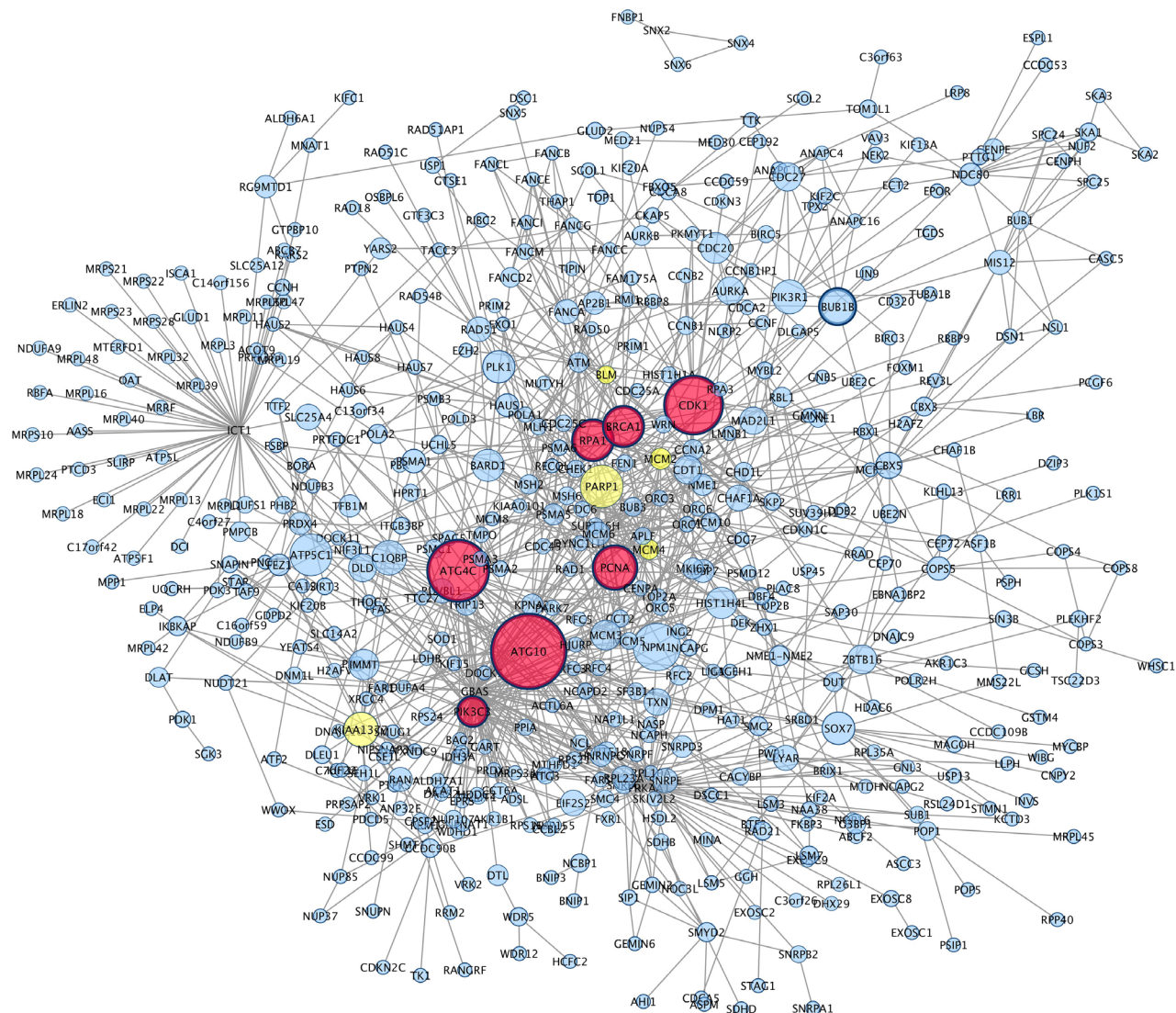


Figure 1: Protein Protein interaction network among up-regulated genes detected in co-cultured keratinocytes and melanocytes from individuals harbouring Red hair color *MC1R* variants (GSE44805 dataset). Nodes represent genes and edges indicate connections between proteins. Nodes are coloured based on number of degree in: low connected node (blue), medium connected node (yellow) and high connected node (red). Node size is proportional to the betweenness centrality value: the higher the value, the larger the node. Hub genes are represented with a thicker black border. The *GBAS*, *PRKAA1*, *ICT1* and *SNX2* hub genes do not follow the node size and border criteria to improve graphical representation of the network.

We detected a gene expression pattern composed of 1,952 transcripts in cutaneous cells from RHC individuals indicating that skin from those individuals presents a distinct gene expression pattern compared to BCH individuals independent of the UV radiation effect. Altogether, these results confirm that loss-of-function *MC1R* variants promote a constitutive genomic deregulation associated with pheomelanin synthesis as observed *in vitro* [8] and in mice models [7]. Three hundred seventy-eight genes were in common in the set of deregulated transcripts both *in vitro* and *in vivo* cutaneous cells from RHC individuals, including a set of the network's hub genes such as *CLN3*, *ATG10*, *WIPI2*, *SNX2*, *PCNA*, *GBAS*, *GABARAPL2* and *YWHAQ*.

Notably, *CLN3*, *ATG10*, *WIPI2*, *SNX2* and *GABARAPL2* are members of the autophagy interaction

network [16]. Autophagy is a highly conserved lysosomal pathway involved in tissue homeostasis, adaptation to starvation and removal of dysfunctional organelles or pathogens [17]. Additional autophagy genes (*ATG12*, *ATG2A*, *ATG9A*, *ATG7*) are also down-deregulated in healthy skin from RHC individuals (Supplementary Table S26). Previous data suggest that regulators of autophagosome formation play a role in melanosome formation and destruction of abnormal melanosomes [18]. Autophagosome formation is a multistep process in which the expansion and closure of the vesicle membrane is controlled by the UBL complex [19]. Notably, most of deregulated genes (*ATG7*, *ATG12*, *ATG10*, *ATG5* and *WIPI*) are related to the UBL complex [20, 21]. Moreover, cutaneous cells from RHC individuals show over-expression of *GABARAPL2* which is essential for

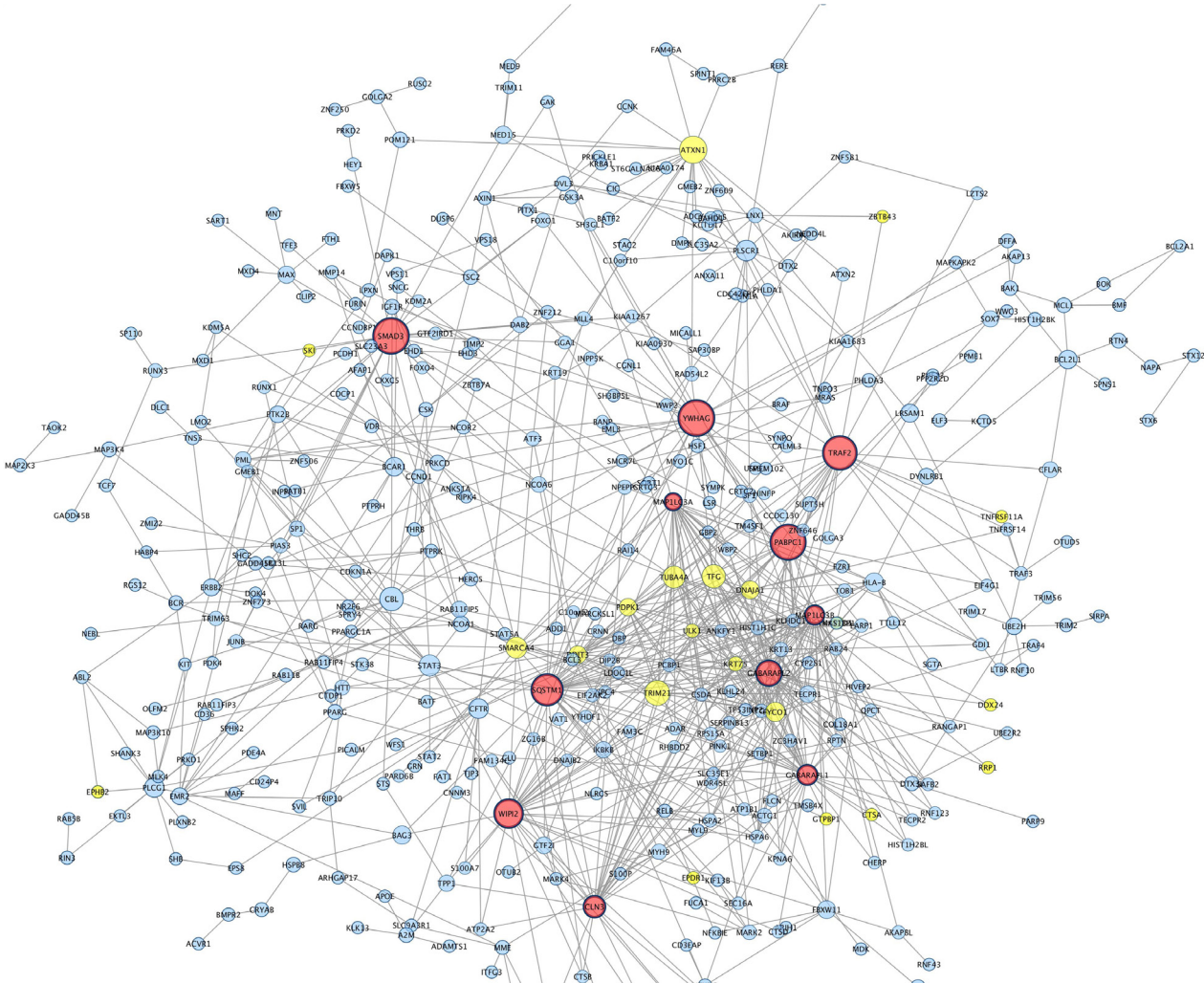


Figure 2: Protein Protein interaction network among down-regulated genes detected in co-cultured keratinocytes and melanocytes from individuals harbouring Red hair color *MC1R* variants (GSE44805 dataset). Nodes represent genes and edges indicate interaction between proteins. Nodes are coloured based on number of degree in: low connected node (blue), medium connected node (yellow) and high connected node (red). Node size indicate Betweenness centrality values. Hub genes are represented with a thicker black border.

Table 1: Topological information of networks detected among deregulated genes detected co-cultured keratinocytes and melanocytes from individuals harbouring Red hair color MC1R variants (GSE44805 dataset)

Degree Centrality						
Network	N. of nodes	N. of edges	Highest Value	Lowest Value	Mean \pm SD	Hub genes*
Up regulated genes	514	1144	77	1	4.45 \pm 6.78	<i>GBAS</i> (77), <i>PRKAA1</i> (54), <i>ICT1</i> (54), <i>ATG4C</i> (42), <i>ATG10</i> (50) <i>PIK3C3</i> (30), <i>PCNA</i> (28), <i>CDK1</i> (27), <i>BRCA1</i> (27) <i>RPA1</i> (25), <i>BUB1B</i> (12) ^a , <i>SNX2</i> (3) ^a <i>GABARAPL2</i> (42) <i>SQSTM1</i> (39) <i>GABARAPL1</i> (38), <i>MAP1LC3B</i> (38), <i>WIPI2</i> (34), <i>MAP1LC3A</i> (31) <i>CLN3</i> (30), <i>YWHAG</i> (29), <i>SMAD3</i> (27), <i>TRAF2</i> (27), <i>PABPC1</i> (24)
Down regulated genes	411	818	42	1	3.98 \pm 6.07	
Betweenness Centrality						
Network	N. of nodes	N. of edges	Highest Value	Lowest Value	Mean \pm SD	Hub genes*
Up regulated genes	514	1144	0.66	0.00	0.01 \pm 0.03	<i>SNX2</i> (0.66), <i>GBAS</i> (0.29), <i>ICT1</i> (0.18), <i>PRKAA1</i> (0.16), <i>ATG10</i> (0.11), <i>ATG4C</i> (0.09), <i>CDK1</i> (0.08), <i>PCNA</i> (0.06), <i>RPA1</i> (0.05), <i>BRCA1</i> (0.05) <i>BUB1B</i> (0.04), <i>PIK3C3</i> (0.03) <i>SMAD3</i> (0.15), <i>YWHAG</i> (0.15), <i>PABPC1</i> (0.14) <i>TRAF2</i> (0.13), <i>SQSTM1</i> (0.12), <i>WIPI2</i> (0.10) <i>CLN3</i> (0.06), <i>GABARAPL1</i> (0.05), <i>MAP1LC3B</i> (0.05) <i>GABARAPL2</i> (0.04), <i>MAP1LC3A</i> (0.04)
Down regulated genes	411	818	0.15	0.00	0.01 \pm 0.02	

Number of nodes, number of edges, Betweenness Centrality and Degree Centrality from both detected networks, *Name of hub genes and its Betweenness centrality and degree values. ^a: *BUB1B* and *SNX2* were selected based on its Betweenness Centrality values.

autophagosome maturation [19] and down-regulation of *CLN3* which play a role in autophagy, endocytosis and vesicular trafficking [16]. Previous comparison of black and red human hair melanosomes reveals differences between eumelanosomes and pheomelanosomes in physical terms such as shape or structural integrity [22]. Our findings support that autophagy genes are closely related to melanosome biogenesis and further suggest that these genes could underlie part of the physical differences between types of melanosomes.

Findings from healthy skins could not be replicated in melanomas. This could be because if the network's hub genes play a role in melanoma development, it may be in early tumor initiation as opposed to progression. In addition, gene expression is under a tighter regulatory control in tumors than in healthy tissues and differences are often subtler, requiring a larger sample to be detected. The largest fold change between RHC and BHC hub gene expression was 1.37 and the power to detect a difference of this magnitude at $\alpha=0.05$ with 20 RHC and

6 BHC tumors was only 37% without multiple testing adjustments, reducing to 5% with adjustment. This is in contrast with the gene showing the largest difference in healthy skins (fold change=2.7, power=0.97 after multiple-tests correction). Nonetheless, the tumor data reinforce the possible involvement of cell cycle homeostasis and autophagy as *PRKAA1* and *PIK3C3* were the top 2 ranked genes. Additional studies of genes identified in this study should be conducted in skin cancer tumors to elucidate their role in malignant transformation.

A previous PPI network analyses conducted in a whole gene expression dataset of 31 primary melanomas and 52 metastases reported a PPI network in which *PCNA*, *CDK1*, *MAD2L1*, *RFC4* and *BRCA1* genes showed highest degree centrality values among upregulated genes

in metastases [14]. We observed a similar behaviour in *in vitro* RHC cutaneous cells since *PCNA*, *CDK1* and *BRCA1* were hub genes in the network of up-regulated genes. However, only *PCNA* reaches statistically significant values in *in vivo* RHC cutaneous cells. Since these genes are essential for cell cycle progression and DNA repair [23–24], these findings suggest that up-regulation of these genes may be caused by an increased and continuous DNA damage. The intrinsic pheomelanin pathway represents an additional contributor to DNA damage by increasing oxidation [7, 8]. However, the molecular process underlying the increased DNA damage is poorly understood. Both *in vitro* and *in vivo* RHC cutaneous cells show an overexpression of *GBAS* which encodes a mitochondrial protein involved in oxidative

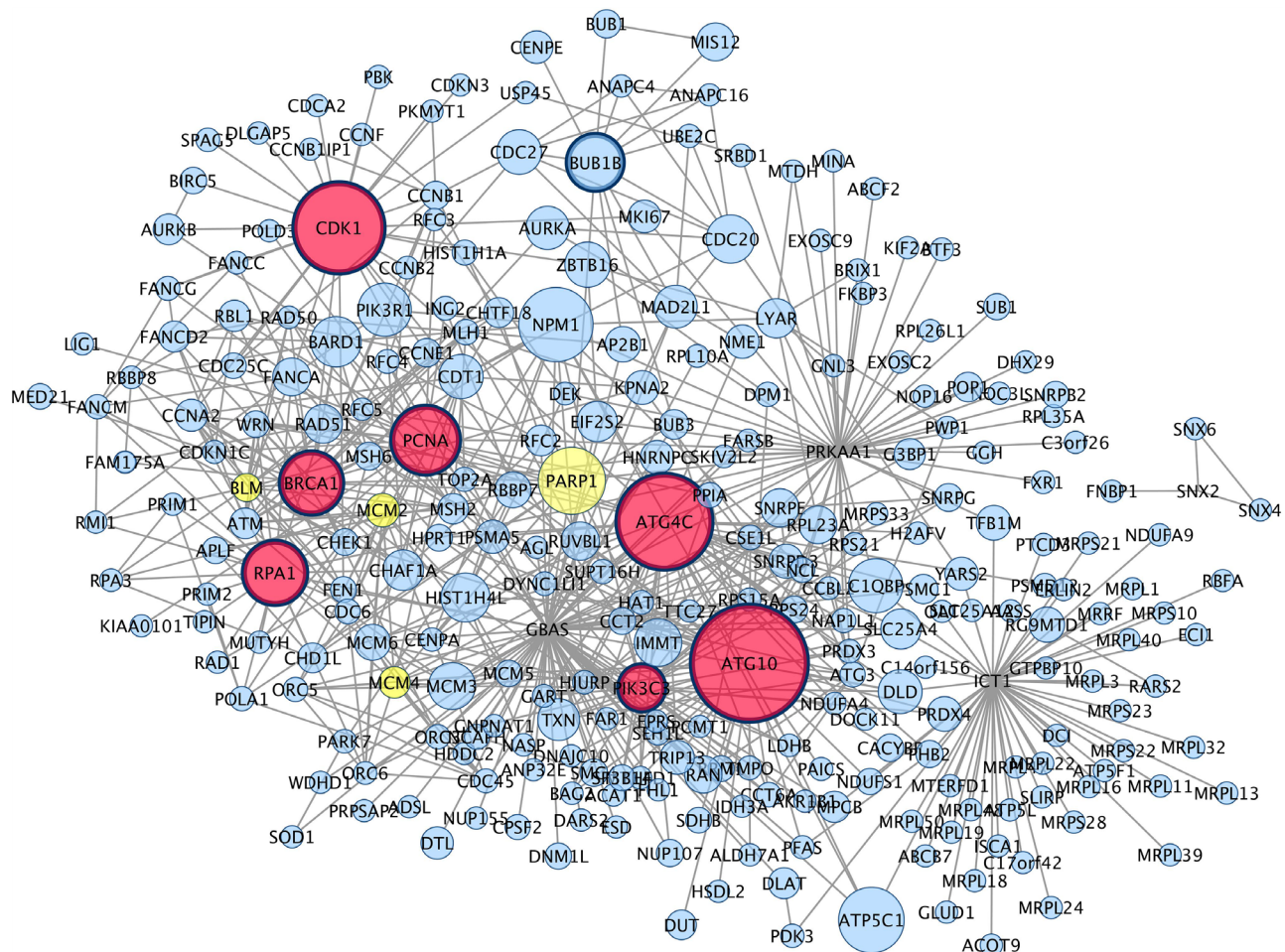


Figure 3: Hub genes and their first connected genes from network detected among up-regulated genes in co-cultured keratinocytes and melanocytes from individuals harbouring Red hair color *MC1R* variants (GSE44805 dataset). Only the first connection of hub genes is indicated. Nodes represent proteins and edges indicate connections between proteins. Nodes are coloured based on number of degree: Red (degrees $> \mu + 3\sigma$; $x > 24.8$), yellow ($\mu + 3\sigma < \text{degree} < \mu + 2\sigma$; $24.8 < \text{degree} < 18.6$), blue (degrees $< \mu + 2\sigma$, degree < 18.6). Node size indicate betweenness centrality values. Hub genes are represented with a thicker black border. The hub genes (N=13 nodes) are directly connected with 270 out of 557 nodes within network. The *GBAS*, *PRKAA1*, *ICT1* and *SNX2* hub genes do not follow the node size and border criteria to improve graphical representation of the network.

phosphorylation [25, 26]. Further studies are required to clarify the biological role of mitochondrial related gene in melanocyte and melanogenesis.

The main limitations of present study are the disparity between platforms used for gene expression capture (resulting in presence/absence of gene probes and probe design), the low frequency of phenotypically

black hair individuals included in the datasets and the differences in type of samples analyzed (cultured skin cells vs non-cultured skin cells) Establishment of cultured cutaneous cells might influence the expression of certain genes resulting in distinct expression such as *ATG10* gene expression which is found up-regulated in cultured cells and down-regulated in non-cultured cells.

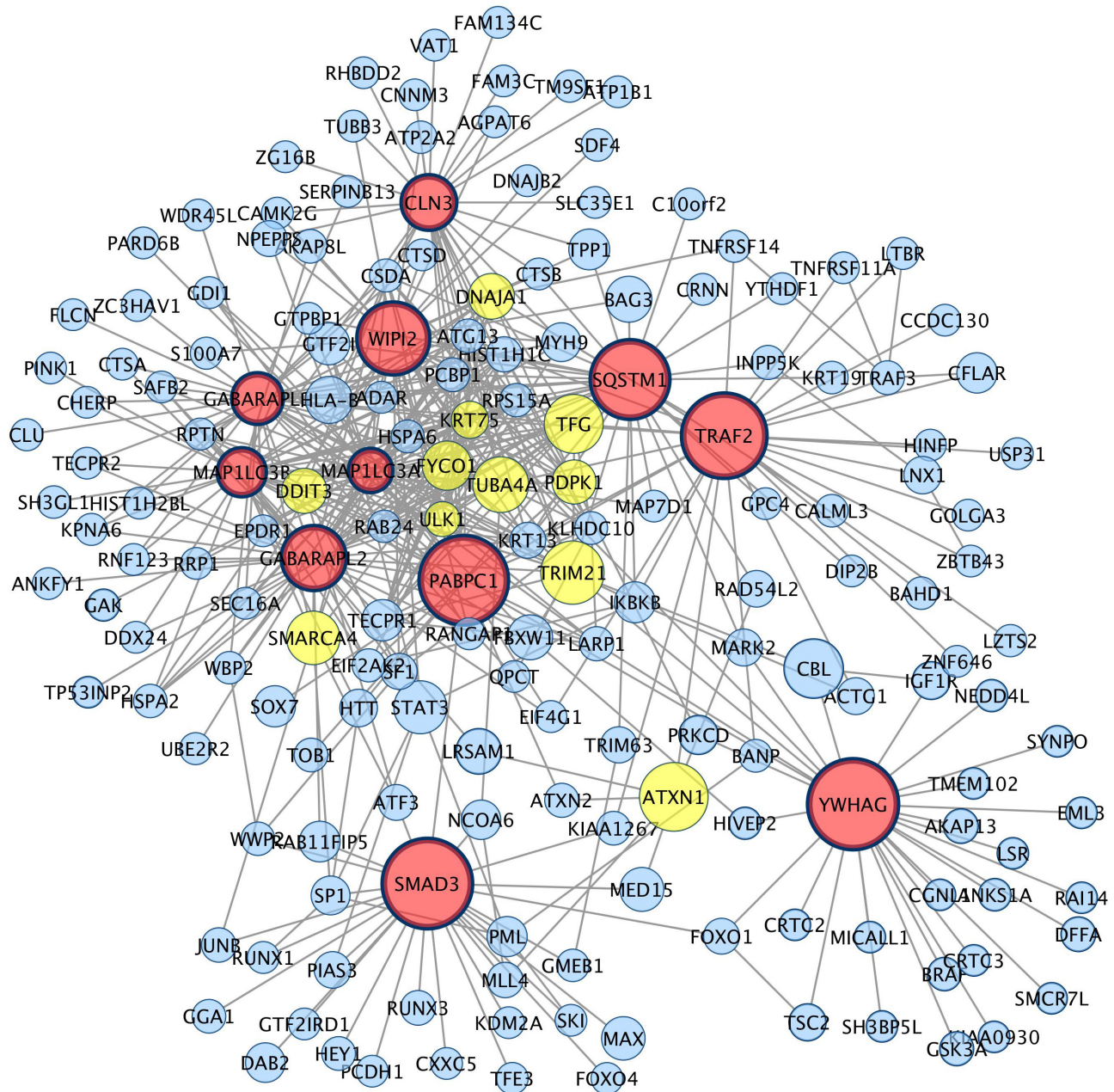


Figure 4: Hub genes and their first connected genes from network detected among down-regulated genes in co-cultured keratinocytes and melanocytes from individuals harbouring Red hair color *MC1R* variants (GSE44805 dataset). Only the first connection of hub genes is indicated. Nodes represent genes and edges indicate direct interaction between proteins. Nodes are coloured based on number of degree in Red (degrees $> \mu + 3\sigma$; $x > 22.2$), yellow ($\mu + 3\sigma < \text{degree} < \mu + 2\sigma$; $22.2 < \text{degree} < 16.3$), blue (degrees $< \mu + 2\sigma$, degree < 16.3). Node size indicate Betweenness centrality values. Hub genes are those nodes with the widest border. The hub genes (N=11 nodes) connects 177 out of 411 nodes.

Table 2: Differential Expression of hub genes in healthy skin from Red hair color individuals vs Black hair color individuals

ID	Results for whole list of genes ^(A) (N=1952)			Results from list of hub genes ^(B) (N=18)		
	logFC	p.value	adj. p.value	logFC	p.value	adj. p.value
<i>ATG10</i>	-0.431	0.001	0.032	-0.432	> 0.001	0.007
<i>CLN3</i>	-0.339	> 0.001	0.024	-0.339	> 0.001	0.007
<i>GABARAPL2</i>	ns	ns	ns	0.254	0.0188	0.056
<i>GBAS</i>	0.374	0.005	0.067	0.373	0.005	0.017
<i>PCNA</i>	0.209	> 0.001	0.026	0.209	0.001	0.007
<i>SNX2</i>	0.506	0.002	0.041	0.508	0.001	0.007
<i>WIPI2</i>	ns	ns	ns	-0.156	0.034	0.076
<i>YWHAG</i>	ns	ns	ns	0.273	0.023	0.060

Differential expression of hub genes was assessed (A) by analyzing the whole gene expression dataset and (B) by analyzing the differential expression of only hub genes. Table shows hub genes with statistically significant p-values and adjusted P-values (adj.p.value). logFC= logarithm of Fold Change; ns=not significant P-value

In summary, the genes identified may help to reveal underlying molecular mechanisms associated with red hair color phenotype and future studies of these genes may provide insight to better understand the increased skin cancer risk observed in individuals harbouring loss-of-function *MC1R* variants.

MATERIALS AND METHODS

Whole genome expression data from three different datasets were included in the study. An initial dataset included whole genome expression from four co-cultured keratinocytes and melanocytes from two pair of siblings. Phenotypical data and genetic variants in melanoma susceptibility genes *MC1R* and *CDKN2A* were obtained in all individuals. A pair of siblings, phenotypically classified as RHC individuals were double heterozygous for the p.R160W, p.R151C variants in *MC1R* (individual 1 and individual 2 were 38 and 33 years old, respectively). In contrast, a pair of siblings with brown hair color carried wild-type *MC1R* alleles (individual 1 and individual 2 were 50 and 51 years old, respectively). Moreover, the germline p.G101W *CDKN2A* mutation was detected in one of each pair of siblings. Extraction of the expression data was carried out using the Whole Human Genome Microarray 4x44K (Agilent). Generation, pre-processing and differential expression and pathway based analyses from this dataset (GSE44805) has been previously reported [8].

Expression profiling in healthy skin tissue was obtained using Illumina Human HT-12 V3 BeadChips (Illumina) in 21 of 705 female individuals from the

TwinsUK Cohort (<http://www.muth.ac.uk/public.html>). The only eligibility criterion for the TwinsUK Cohort was twin status and therefore the sample is held to be representative of the general population [27, 28]. For the present study, individuals were selected based on phenotypical data since the genetic status of the *MC1R* gene was not available. Fourteen red hair individuals (two monozygotic twin pairs, one dizygotic twin pair, and 8 singletons) and 7 with black hair (all singletons) of Caucasian ancestry were selected. The mean age for the RHC individuals was 50.88 years old (range=43.33-59.84), while for the BHC individuals was 62.04 years old (range=49.23-74.54). All red hair individuals but two (missing the *MC1R* genotype) carry at least one RHC allele. For the pair without genetic data, multiple photographs taken in different years were used in order to confirm that they actually had red hair. Non RHC alleles were detected in black hair individuals. Probe expression levels were log transformed and quantile normalized, and only probes mapping uniquely to genes of known function and not containing common SNPs were included in the analysis, as detailed elsewhere [28], resulting in 16,646 genes. Mean expression profile was used when the same gene was targeted by multiple probes. Written informed consent was provided by all the twins, and the Guy's and St Thomas' Hospital NHS Trust Research Ethics Committee approved the study.

Generation, pre-processing and analysis of whole genome gene expression in melanoma tumors from the Leeds Melanoma Cohort (LMC, ethical approval MREC 1/3/57, PIAG 3-09(d)/2003) has been described elsewhere [29]. This data contains 204 melanoma primaries

expression-profiled using DASL array HT12 v4. Gene expression was compared between 20 melanoma tumors from RHC and 6 melanoma tumors from BHC patients. All RHC melanoma patients carried at least one RHC allele). One patient has missing MC1R genotype. We observed that among black haired individuals, one patient carried two RHC variants. One patient was missing their MC1R genotype.

Statistical analyses

Standardization of gene expression data across arrays from co-cultured keratinocytes and melanocytes was performed with Agilent Processed Signal (Agilent Feature Extraction Software) using quantile normalization [30]. Differential gene expression analysis was carried out using the *limma* [31] package from Bioconductor. Multiple testing adjustment of p-values was performed according to Benjamini and Hochberg (Benjamini & Hochberg, 1995; Benjamini and Yekutieli, 2001) controlling False Discovery Rate (FDR = 5%).

Protein-protein interaction (PPI) networks were built based on differentially expressed genes. Identification of statistically significant PPI networks within the set of differentially expressed genes was carried out by SNOW [32] as implemented in Babelomics 5 suite [33], where approaches for protein-protein interaction data from 5 databases such as IntAct Molecular Interaction Database [34], Molecular INTeraction database [35], Biomolecular Interaction Network Database [36], Database of Interacting Proteins [37] and Human Protein Reference database [38] are included for functional profiling of genomic data, evaluating the cooperative behavior of a list of genes as a functional module. SNOW calculates the Minimal connected network (MCN) for the proteins/genes in the list. The topology of this MCN is then compared against 10,000 random MCNs with same size range to obtain a p-value. Statistically significant connected networks of differentially expressed genes were obtained for up and down regulated transcripts in RHC co-cultured melanocytes and keratinocytes. Further visualization and evaluation of networks was conducted using CytoscapeWeb [39]. Networks were evaluated using two topological parameters, namely degree and betweenness centrality. Degree centrality counts the number of edges connected to each node (protein/gene). Betweenness centrality counts the number of times each node is included in the shortest path between any two other nodes, indicating which node(s) have the larger influence on the communication between proteins/genes included in the network. Nodes were classified based on the mean (μ) and standard deviation (σ) of degree centrality value calculated over all proteins/genes in: lower connected node ($\text{degree} < \mu + 2\sigma$), medium connected node ($\mu + 2\sigma > \text{degree} > \mu + 3\sigma$), and highly connected node ($\text{degree} > \mu + 3\sigma$). Highly connected nodes or those not

highly connected nodes but with high betweenness centrality values were classify as top nodes (hub genes).

Differential gene expression analysis in healthy skin tissue was carried out using the *limma* R package controlling for family structure and batch effects. Benjamini and Hochberg's method, as implemented in *limma*, was used to control the false discovery rate at 10%. Differential expression of the hub genes identified in the gene expression data from co-cultured melanocytes and keratinocytes was assessed firstly by analysing the whole gene expression dataset and then by restricting it to hub genes only.

Whole differential gene expression was evaluated by SNOW following the same analyses as in co-cultured melanocytes-keratinocytes data.

Differential expression analysis in tumors was restricted to the genes that were identified as hub genes in the melanocytes-keratinocytes dataset (N=23). Examination of the probe list in the Illumina DASL whole genome array confirmed that 33 probes were arrayed covering all 23 genes selected as hub genes in networks. Differential expression in tumors from 20 RHC vs 6 BHC patients from LMC was performed using linear regression and adjusting for potential batch effect. Analysis was conducted using STATA v12 software.

CONFLICTS OF INTEREST

There is no conflicts of interest to disclose.

GRANT SUPPORT

This work was mainly supported by intramural grants INTRA/07/726.2 and ACCI2014 (U726 and U714) from the Biomedical Network Research on Rare Diseases (CIBERER) of the Instituto de Salud Carlos III, Spain. The research at the Melanoma Unit in Barcelona is partially funded by Spanish Fondo de Investigaciones Sanitarias (PI15/00716 and grants PI15/00956); AGAUR 2014 SGR 603 of the Catalan Government, Spain; European Commission under the 6th Framework Programme, Contract No.LSHC-CT-2006-018702 (GenoMEL) and by the European Commission under the 7th Framework Programme, Diagnostix; The National Cancer Institute (NCI) of the US National Institute of Health (NIH) (CA83115) and a grant from "Fundació La Marató de TV3, 201331-30", Catalonia, Spain. The Department of Bioengineering UC3M-CIEMAT-IISFJD in Madrid is partially funded from the Spanish Ministerio de Economía y Competitividad (SKINNOVA; SAF2013-43475-R) and (CELLCAM; S2010/BMD-2420) from Comunidad de Madrid. The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The

study also received support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. MF, VB and AV are supported by BSF grant n. 5044i. The Leeds contribution (tumour gene expressions) was funded by Cancer Research UK Programme grant C588/A19167 and Project Grant C8216/A6129 and NIH Award CA83115.

REFERENCES

1. Abdel-Malek ZA, Swope VB, Starner RJ, Koikov L, Cassidy P, Leachman S. Melanocortins and the melanocortin 1 receptor, moving translationally towards melanoma prevention. *Arch Biochem Biophys*. 2014; 563:4-12.
2. Garcia-Borron JC, Abdel-Malek Z, Jimenez-Cervantes C. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. *Pigment Cell Melanoma Res*. 2014; 27:699-720.
3. Pasquali E, Garcia-Borron JC, Fargnoli MC, Gandini S, Maisonneuve P, Bagnardi V, Specchia C, Liu F, Kayser M, Nijsten T, Nagore E, Kumar R, Hansson J, Kanetsky PA, Ghiorzo P, Debnik T, et al. MC1R variants increased the risk of sporadic cutaneous melanoma in darker-pigmented Caucasians: a pooled-analysis from the M-SKIP project. *Int J Cancer*. 2015; 136:618-631.
4. Tagliabue E, Fargnoli MC, Gandini S, Maisonneuve P, Liu F, Kayser M, Nijsten T, Han J, Kumar R, Gruis NA, Ferrucci L, Branicki W, Dwyer T, Blizzard L, Helsing P, Autier P, et al. MC1R gene variants and non-melanoma skin cancer: a pooled-analysis from the M-SKIP project. *Br J Cancer*. 2015; 113:354-363.
5. Rouzaud F, Kadekaro AL, Abdel-Malek ZA, Hearing VJ. MC1R and the response of melanocytes to ultraviolet radiation. *Mutat Res*. 2005; 571:133-152.
6. Hill HZ, Hill GJ. UVA, pheomelanin and the carcinogenesis of melanoma. *Pigment Cell Res*. 2000; 13:140-144.
7. Mitra D, Luo X, Morgan A, Wang J, Hoang MP, Lo J, Guerrero CR, Lennerz JK, Mihm MC, Wargo JA, Robinson KC, Devi SP, Vanover JC, D'Orazio JA, McMahon M, Bosenberg MW, et al. An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature*. 2012; 491:449-453.
8. Puig-Butille JA, Escamez MJ, Garcia-Garcia F, Tell-Marti G, Fabra A, Martinez-Santamaria L, Badenas C, Aguilera P, Pevida M, Dopazo J, del Rio M, Puig S. Capturing the biological impact of CDKN2A and MC1R genes as an early predisposing event in melanoma and non melanoma skin cancer. *Oncotarget*. 2014; 5:1439-1451.
9. Yue Z, Li HT, Yang Y, Hussain S, Zheng CH, Xia J, Chen Y. Identification of breast cancer candidate genes using gene co-expression and protein-protein interaction information. *Oncotarget*. 2016.
10. Keller MP, Choi Y, Wang P, Davis DB, Rabaglia ME, Oler AT, Stapleton DS, Argmann C, Schueler KL, Edwards S, Steinberg HA, Chaibub Neto E, Kleinhanz R, Turner S, Hellerstein MK, Schadt EE, et al. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res*. 2008; 18:706-716.
11. Deng SP, Zhu L, Huang DS. Predicting Hub Genes Associated with Cervical Cancer through Gene Co-Expression Networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2016; 13:27-35.
12. Diao H, Li X, Hu S, Liu Y. Gene expression profiling combined with bioinformatics analysis identify biomarkers for Parkinson disease. *PLoS One*. 2012; 7:e52319.
13. Sun H. Identification of key genes associated with gastric cancer based on DNA microarray data. *Oncol Lett*. 2016; 11:525-530.
14. Qiu T, Wang H, Wang Y, Zhang Y, Hui Q, Tao K. Identification of genes associated with melanoma metastasis. *Kaohsiung J Med Sci*. 2015; 31:553-561.
15. Almaas E. Biological impacts and context of network theory. *J Exp Biol*. 2007; 210:1548-1558.
16. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. *Nature*. 2010; 466:68-76.
17. Levine B, Kroemer G. Autophagy in the pathogenesis of disease. *Cell*. 2008; 132:27-42.
18. Ho H, Ganesan AK. The pleiotropic roles of autophagy regulators in melanogenesis. *Pigment Cell Melanoma Res*. 2011; 24:595-604.
19. Tanida I. Autophagy basics. *Microbiol Immunol*. 2011; 55:1-11.
20. Nemoto T, Tanida I, Tanida-Miyake E, Minematsu-Ikeguchi N, Yokota M, Ohsumi M, Ueno T, Kominami E. The mouse APG10 homologue, an E2-like enzyme for Apg12p conjugation, facilitates MAP-LC3 modification. *J Biol Chem*. 2003; 278:39517-39526.
21. Dooley HC, Razi M, Polson HE, Girardin SE, Wilson MI, Tooze SA. WIPI2 links LC3 conjugation with PI3P, autophagosome formation, and pathogen clearance by recruiting Atg12-5-16L1. *Mol Cell*. 2014; 55:238-252.
22. Liu Y, Hong L, Wakamatsu K, Ito S, Adhyaru B, Cheng CY, Bowers CR, Simon JD. Comparison of structural and chemical properties of black and red human hair melanosomes. *Photochem Photobiol*. 2005; 81:135-144.
23. Moldovan GL, Pfander B, Jentsch S. PCNA, the maestro of the replication fork. *Cell*. 2007; 129:665-679.
24. Tanaka S, Tak YS, Araki H. The role of CDK in the initiation step of DNA replication in eukaryotes. *Cell Div*. 2007; 2:16.
25. Smits P, Rodenburg RJ, Smeitink JA, van den Heuvel LP. Sequence variants in four candidate genes (NIPSNAP1, GBAS, CHCHD1 and METT11D1) in patients with

- combined oxidative phosphorylation system deficiencies. *J Inherit Metab Dis*. 2010; 33:S13-19.
26. Martherus RS, Sluiter W, Timmer ED, VanHerle SJ, Smeets HJ, Ayoubi TA. Functional annotation of heart enriched mitochondrial genes GBAS and CHCHD10 through guilt by association. *Biochem Biophys Res Commun*. 2010; 402:203-208.
 27. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res*. 2001; 4:464-477.
 28. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin SY, Glass D, Travers M, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012; 44:1084-1089.
 29. Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, Wolter P, Patel P, Sucker A, Schadendorf D, Jonsson GB, Bishop DT, Newton-Bishop J. Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. *Oncotarget*. 2015; 6:11683-11693.
 30. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003; 19:185-193.
 31. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3:Article3.
 32. Minguéz P, Gotz S, Montaner D, Al-Shahrour F, Dopazo J. SNOW, a web-based tool for the statistical analysis of protein-protein interaction networks. *Nucleic Acids Res*. 2009; 37:W109-114.
 33. Alonso R, Salavert F, Garcia-Garcia F, Carbonell-Caballero J, Bleda M, Garcia-Alonso L, Sanchis-Juan A, Perez-Gil D, Marin-Garcia P, Sanchez R, Cubuk C, Hidalgo MR, Amadoz A, Hernansaiz-Ballesteros RD, Aleman A, Tarraga J, et al. Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res*. 2015; 43:W117-121.
 34. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012; 40:D841-846.
 35. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardoza AP, Santonico E, Castagnoli L, Cesareni G. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*. 2012; 40:D857-861.
 36. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res*. 2003; 31:248-250.
 37. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*. 2004; 32:D449-451.
 38. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, et al. Human Protein Reference Database--2009 update. *Nucleic Acids Res*. 2009; 37:D767-772.
 39. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. Cytoscape Web: an interactive web-based network browser. *Bioinformatics*. 2010; 26:2347-2348.